Reasoning About Relations: Spatial and Nonspatial Problems

Manuel Carreiras and Carlos Santamaría

Universidad de La Laguna, Tenerife, Spain

Two experiments investigated the mental representation of spatial and nonspatial two-dimensional problems. The experiments were designed to contrast opposite predictions of the model theory of reasoning and the formal rules of inference theories. Half of the problems required more inferential steps but only one model, whereas the other half required fewer inferential steps but two models. According to the inference rules, theory problems that require more inferential steps should be harder, whereas the model-based theory predicts that problems that require two models should be harder. In Experiments 1a and 1b we measured the problem solving time and the percentage of errors. In Experiments 2a and 2b the problems were presented segmented in two different displays. We measured the comprehension time for each display, the question answering times, and the percentage of errors. The results of all experiments supported the model theory predictions in both spatial and nonspatial domains.

INTRODUCTION

Some of the most frequent everyday inferences involve the deduction of a relationship among a set of entities from a known relation of these entities with other elements. For example, suppose that someone is ignorant of the height of the new Monument X and is told:

Monument X is higher than Big Ben
The Eiffel Tower is higher than Monument X

He or she could infer that the height of Monument X is somewhere between that of Big Ben and the Eiffel Tower, but also that Big Ben is the shortest and the Eiffel Tower the highest of these three monuments.
The nature of the cognitive mechanisms underlying the resolution of such problems has been controversial since the 1960s. Some authors considered that to solve these problems people build a sort of spatial representation of the situation described in the problem. This spatial model was supported by DeSoto, London, and Handel (1965), and Huttenlocher and her colleagues (Huttenlocher, 1968; Huttenlocher & Higgins, 1971), who stressed the importance of the semantics of the situation rather than the syntactics of the phrasing. For instance, a pair of premises like:

\[ A > B \]
\[ B > C \]
could be mentally represented as:

\[ A \]
\[ B \]
\[ C \]

According to this position, the success in finding the right solution to these kind of problems depends on the difficulties of building and scanning such a representation.

An alternative theory was proposed by Clark (1969) and Bar-Hillel (1967), who claimed that participants build a language-like mental representation of the premises in order to solve such problems. The conclusion is drawn by applying formal rules of inference to the linguistic representation. For example, a transitive relation like “bigger than” yields the application of the transitivity rule:

For any \( x, y, z \),
\[ \text{if } x \text{ is bigger than } y \text{ and } y \text{ is bigger than } z, \]
\[ \text{then } x \text{ is bigger than } z \]

As Byrne and Johnson-Laird (1989) have pointed out, the kind of problems used at that time for research on this topic prevented experimenters from conducting a crucial experiment. These problems, called three-term series problems or linear syllogisms, are easy to solve and so simple to explain that both types of theories (spatial and linguistic) produce similar predictions.

Important improvements have occurred recently in this area, both in methodology and theory. A crucial methodological enhancement was the use of spatial descriptions. Using spatial descriptions Mani and Johnson-Laird (1982) demonstrated the existence of two forms of representation: linguistic representations, which are close to the surface form of the text, and discourse models or mental models, representations of what the text is about, whose
structure parallels the structure of the part of the world it represents, and not
the structure of the text in which it is described. In particular, Mani and
Johnson-Laird (1982) have shown that determinate passages consistent with
only a single model or spatial layout are easier to remember than indeterminate
descriptions that are consistent with more than one layout. They presented to
the participants determinate spatial descriptions such as:

1. The spoon is to the left of the knife
2. The plate is to the right of the knife
3. The fork is in front of the spoon
4. The cup is in front of the knife

which is consistent with the following layout of objects:

spoon   knife   plate
fork     cup

They also presented indeterminate spatial descriptions to the participants,
which were constructed from the determinate ones by changing the last word
of the second sentence. The indeterminate descriptions are consistent with two
different layouts;

1. The spoon is to the left of the knife
2. The plate is to the right of the spoon.
3. The fork is in front of the spoon
4. The cup is in front of the knife

spoon   knife   plate  spoon   plate   knife
fork     cup     fork   cup

The participants’ task was to listen to a series of determinate and indeterminate
descriptions and classify them as true or false with respect to different layouts
presented to them. Then, they were given an unexpected recognition test of
their memory for the descriptions. Participants remembered the gist of the
determinate descriptions better than the gist of the indeterminate descriptions,
but they remembered the verbatim detail of indeterminate descriptions better
than that of determinate descriptions. Such results were interpreted as
suggesting that participants constructed a mental model of determinate
descriptions to compare with layouts, which preserved the structure of the
layout but did not preserve the verbatim detail. However, participants did not
construct a mental model of indeterminate descriptions but a surface
representation closer to the linguistic structure of the descriptions. The
existence of two forms of representation—linguistic representation and a
mental mode—has been extended and corroborated in a number of other experiments (e.g. Ehrlich & Johnson-Laird, 1982; Garnham, 1987). As to the theoretical underpinnings, the improvements were even more dramatic. The analogical (image) model is one of the precedents of the mental-model theory of reasoning developed recently (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). For this theory, the main cause of difficulty in reasoning problems is the number of different representations of the state of affairs described in the problem (models) that should be constructed in order to solve the problem. The linguistic theory was a source for the contemporary proof theoretic theories (Braine, 1990; Hagert, 1984; Rips, 1983, 1994). For these theories the difficulty of a problem is mainly determined by the number of inferential steps in the derivation.

The use of four-term spatial problems allows a comparison of the two kinds of theories. For instance, problems like (I) and (II), that follow, lead to divergent predictions from the two sorts of theories.

Problem I

1. A is on the right of B
2. C is on the left of B
3. D is in front of C
4. E is in front of A.

Hence, D is on the left of E.

```
    C  B  A
D   E
```

Problem II

1. A is on the right of B
2. C is on the left of A
3. D is in front of C
4. E is in front of A

Hence, D is on the left of E.

```
    C  B  A  B  C  A
D  E  D  E
```

An instantiation of a inference-rule system for two-dimensional problems was described by Hagert (1984). According to this model, Problem I needs more derivational work than Problem II, because there is no single premise
expressing the relation between the pair of items that $D$ and $E$ are in front of. That is, Problem II includes a premise stating that $C$ is on the left of $A$, but this has to be inferred in Problem I. Consequently, if people are using such rules, Problem I should be more difficult to solve than Problem II. Inference-rule theory predicts that problems that require more steps in their formal derivation to reach the conclusion will be harder. In contrast, the model theory predicts that Problem I should be easier to solve than Problem II, because the premises of Problem I yield only one model whereas those of Problem II yield two models. The model theory predicts that a problem that requires only one model will be easier than one that requires more than one model.

Byrne and Johnson-Laird (1989) investigated whether the number of inferential steps or the number of models predicted differences in the difficulty of spatial problems. They asked participants to listen to spatial descriptions similar to Problems I and II mentioned previously, and then to judge the relative location of two objects. In their first experiment, in which they held constant the number of inferential steps but varied the number of models, they found a higher percentage of correct responses for one-model problems than for multiple-model problems. In their second experiment, they pitted the inference-rule and model-based theories against each other on problems where the two theories predicted opposite outcomes. In particular, they contrasted one-model problems and multiple-model problems, where the one-model problems required more inferential steps than multiple-model problems. They found that problems that require more models were harder than those that require only one model, but problems that require more inferential steps were not harder than those that required fewer steps. Therefore, their results supported the predictions of the model-based theory.

As in most studies of reasoning, Byrne and Johnson-Laird (1989) relied on errors in performance to make inferences about how spatial descriptions are encoded, represented, and solved. Most of the previous research has mainly focused on the end-product of the reasoning process. However, reasoning based on mental models involves an on-line comprehension process, which implies an incremental updating of the representation on the basis of the present and past input. People try to build an integrated representation of the discourse piece by piece as they read or hear it, so the information incorporated into the representation is continuously updated by the incoming information. In that sense, a mental model provides both representations of what the discourse is about and at the same time a representation for understanding the following sentence; that is, the resulting representation at any given moment guides the interpretation of subsequent input. As mental models are constructed in an incremental way while reading and understanding the premises, it is very important to capture their time course. Looking at how participants develop representations from spatial descriptions will help us to contrast predictions from inference-rule and model-based theories.
In this paper we will consider how well participants integrate information from different premises and how well they are able to develop representations from verbal descriptions by measuring their reading times. In our first experiment we measure the time participants take to comprehend the problems and to solve them. Problem solving time is a very valuable convergent measure. If participants solve problems by applying derivation rules, they should spend longer reading descriptions that require more derivation rules. On the contrary, if participants integrate the premises into a mental model, then they should spend longer reading descriptions consistent with more than one model as compared to those that are only consistent with one model. In our second experiment the problems are presented segmented into several displays. We measure the comprehension time for each display and the question answering times. This is a more fine-grained measure that allows us to capture the updating process of the mental representations, how well participants integrate information from different parts of the discourse, and what mental operations—inferential rules vs. mental models—are involved in finding the solution to the problems. In both experiments the inference-rule and the model-based theories make different predictions.

Another important improvement of this investigation is that both spatial and nonspatial descriptions are investigated. The spatial description task has been proved successful in providing evidence for deciding between the two hypotheses. However, it could be argued that mental models are only applicable to the spatial domain (Rips, 1994). In the present research we test predictions of the inference-rule and model-based hypotheses using spatial and nonspatial descriptions. If participants use mental models or inference rules as their usual reasoning strategy, we should obtain the same pattern of data with spatial and nonspatial problems. As stated before, according to the inference-rule hypothesis, problems that require more steps for their derivation should be harder, whereas according to the model-based hypothesis, problems that require more models should be harder. On the other hand, if participants used mental models for spatial reasoning but inference rules for non-spatial reasoning then a different pattern of data should arise for the two types of problems. Finally, another possibility is that nonspatial problems are solved by analogy to spatial ones. They are translated first to the spatial domain and then resolved as if they were spatial. If this were the case we should observe that participants should take longer to solve nonspatial problems as compared to spatial ones, but show the same basic pattern of data.

**EXPERIMENT 1A**

The purpose of the first experiment was to investigate whether the inference-rule hypothesis or the model-based hypothesis, which generate opposite predictions concerning the two-dimensional problems, better accounts for how
people reason about spatial relations among objects. According to the model-based theory, one-model problems (I) should be easier to solve than two-model problems (II). On the contrary, the inference-rule theory predicts that problems that require fewer inferential steps (II) should be easier than problems that require more steps to be solved (I). The difficulty of the problems is measured by the time participants take to solve the different problems, as well as by the percentage of errors.

**Method**

*Participants.* The participants were 29 undergraduate students from the University of La Laguna who received course credit for their participation.

*Design and Materials.* Eight experimental problems were constructed. Four of them were one-model problems (Problem I) and the other four were two-model problems (Problem II). For instance:

**Problem I**

1. A is on the right of B
2. C is on the left of B
3. D is in front of C
4. E is in front of A

Who is more to the left D or E?

**Problem II**

1. A is on the right of B
2. C is on the left of A
3. D is in front of C
4. E is in front of A

Who is more to the left D or E?

Each problem was presented with different proper names. For each problem, five different familiar proper names of approximately the same length (all of them were two syllables long, with five or six 11 letters) that began with different initial letters were selected. The eight experimental problems were always presented in a horizontal spatial orientation, as in the examples shown here. There were four different horizontal orientations for each type of problem, resulting from the order in which the premises were presented (1,2,3,4; 1,2,4,3; 2,1,3,4; and 2,1,4,3). There were also eight filler problems.
They were included to reduce the proportion of problems presented with the same spatial orientation, so that participants would not get used to the same type of orientation. These problems had different numbers of elements from the experimental problems and/or repeated relational terms, so they were not analysed.

The problems were presented in a different random order to each participant. In addition, in half of the trials the correct answer corresponded to the first-mentioned proper name in the question, whereas in the other half it corresponded to the second-mentioned proper name.

Procedure. Participants were tested individually in a small quiet room. Each participant read the eight experimental problems intermixed randomly with each other and with eight other problems that acted as fillers to divert attention from the structure of the materials.

The experiment was controlled on-line by a IBM compatible computer. The participants’ task was to read the problems and the questions at their own pace, and to answer the questions as quickly and accurately as possible. For each problem the four premises and the question, which contained two proper names, were presented together in the centre of the screen. Participants had to press the return key when they were ready to write the solution for the problem. The computer recorded participants’ reading times of the premises and the questions. After pressing the return key, which caused the problem to disappear from the screen, they had to use the keyboard to write the solution of the problem, which was one of the proper names that appeared in the question. Before each problem the prompt:

   PULSE LA BARRA ESPACIADORA PARA CONTINUAR
   [Press the space bar to continue]

appeared on the screen. When it was presented, participants had to press the space bar to display the premises and the question corresponding to each problem. The instructions stressed that the problems were to be answered as quickly and as accurately as possible, but emphasising accuracy over speed.

Before the presentation of the experimental materials there were three practice trials whose primary purpose was to familiarise participants with the procedure.

Results and Discussion

Table 1 shows the mean response times and the percentage of correct responses for one- and two-model problems. (The error response times are discarded from the analyses and tables throughout this paper.) Overall the percentage of correct responses was high for both types of problems. However, the percentage
of correct responses for the one-model problems was reliably higher than for the two-model problems (Wilcoxon’s T; Z=2.59, N= 29, P < .01). In addition, one-model problems were solved faster than two-model problems [t(28)=3.51, P < .002]. Both the percentage of errors and the correct response times support the model-based predictions. Participants were faster and more accurate in one-model problems than in two-model problems. This result cannot be accounted for by the inference rules hypothesis, which predicted the opposite result.

EXPERIMENT 1B

The goal of this experiment was to investigate if the results obtained in Experiment 1a could be extended to a nonspatial domain. The same hypotheses as those of Experiment 1a were formulated for a different set of nonspatial descriptions; that is, two-dimensional problems that described a different type of relations among characters (nonspatial relations).

Method

Participants. The participants were 29 undergraduate students from the same population as those tested in Experiment 1a. They received course credit for their participation. None of them had taken part in the previous experiment.

Design and Materials. Another set of materials with the same structure to those of Experiment 1a but with a different content were constructed. Instead of spatial terms, the verbs study (more/less) and copy from were used in the problems of the present experiment. Four of them were one-model problems

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Experiments 1a and 1b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One model (I)</td>
</tr>
<tr>
<td>Spatial problems</td>
<td></td>
</tr>
<tr>
<td>response times</td>
<td>21.7 (9.3)</td>
</tr>
<tr>
<td>correct responses</td>
<td>99</td>
</tr>
<tr>
<td>Nonspatial problems</td>
<td></td>
</tr>
<tr>
<td>response times</td>
<td>17.9 (7.7)</td>
</tr>
<tr>
<td>correct responses</td>
<td>97</td>
</tr>
</tbody>
</table>

Mean correct response times (in seconds), standard deviations (parentheses), and percentage of correct responses to the Problems of Experiments 1a and 1b
(Problem I) and the other four were two-model problems (Problem II). For instance:

**Problem I**
1. A studied more than B
2. C studied less than B
3. D copied from A
4. E copied from C
Who got better marks D or E?

**Problem II**
1. A studied more than B
2. C studied less than A
3. D copied from A
4. E copied from C
Who got better marks D or E?

Apart from the fact that the problems had a different content (nonspatial relations), the design and materials were similar to those of Experiment 1a.

**Procedure.** The procedure was the same as that of Experiment 1a.

**Results and Discussion**

The mean correct response times and the percentage of correct responses for one- and two-model problems are shown in the bottom part of Table 1. The percentage of correct responses was very high for both types of problems. Because performance was close to ceiling, no differences in percentage of correct responses were found between both types of problems (Wilcoxon’s T; $Z=0.53$, $N=29$, $P<1$). As for the response times, one-model problems were solved faster than the two-model problems [$t(28)=4.44$, $P<.0001$]. Thus, these response times replicate those obtained in Experiment 1a and support the model-based hypothesis, but not the inference-rule hypothesis, which predicted the opposite results. They also suggest that participants follow a similar strategy to resolve both spatial and nonspatial problems. So, reasoning by mental models does not seem to be restricted to the spatial domain. However, the lack of differences in accuracy does not replicate what we have found in the previous experiment. As we pointed out, the very high performance of the participants in this task caused a ceiling effect that prevented differences from arising.
EXPERIMENT 2A

In this experiment we again tested predictions from the inference-rule hypothesis and the model-based hypothesis in relation to the processing load imposed by parts of the spatial descriptions. By presenting the spatial descriptions segmented into several displays we can measure the relative processing difficulty of several pieces of text and contrast them with specific predictions from the two hypotheses. Moreover, the segmented presentation of the spatial descriptions also allows us to examine how participants integrate information from different parts of the text and how updating takes place.

The spatial descriptions were presented in three different displays. The first display contained premises 1 and 2, the second display contained premises 3 and 4, and the third display the question. According to the mental models hypothesis, the critical difference between Problem I and Problem II will occur in the first display. When reading the first display participants will place a foundation for one-model representation in Problem I while they will establish a framework for a two-model representation in Problem II. Therefore, participants should spend more time reading the first display (premises 1 and 2) of Problems II than of Problems I. On the contrary, the inference-rule hypothesis predicts that participants should spend more time reading the first display of Problem I than of Problem II. As has been shown in the description of derivational steps for two-dimensional problems, there is no premise declaring the explicit relation between C and A in Problem I, so participants need to perform this additional computation. However, this additional computation is not needed in Problem II. In addition, as in the previous experiments, the model-based hypothesis predicts that overall one-model problems (I) should be easier than two-model problems (II), while the inference-rule hypothesis makes the opposite prediction. Problems (II) that require more inferential steps to come to the conclusion will be harder than the problems of type I. Therefore, besides the predicted differences in the reading times of the first display, both hypotheses also predict differences in the percentage of errors and/or in the question answering times.

Method

Participants. Thirty-one undergraduate students from the University of La Laguna participated in this experiment in fulfilment of a course credit. None of them had participated in the previous experiments.

Design and Materials. The design and materials were the same as those of Experiment 1a.
Procedure. As in Experiments 1a and 1b each participant read the eight experimental problems intermixed randomly with each other and with eight other problems which acted as fillers to divert attention from the structure of the materials. The only difference between the procedure of the present experiment and that of the previous experiments was that each problem was presented in several displays.

The experiment was controlled on-line by a IBM compatible computer. The participants’ task was to read the problems and the questions at their own pace, and to answer the questions as quickly and accurately as possible. Each problem was presented in three successive displays. The first display contained premises 1 and 2, the second display premises 3 and 4, and the third display the question. Before each problem the prompt:

PULSE LA BARRA ESPACIADORA PARA CONTINUAR

[Press the space bar to continue]

appeared on the screen. When it was presented, participants had to press the space bar to display the first two premises of a problem. A further press caused the first two sentences to disappear and the last two premises to be displayed. A third key press erased the last two premises from the screen, and they were replaced by the question. The question contained two proper names. Participants had to press the return key when they were ready to write the solution of the problem. After pressing the return key, the screen went blank, and participants had to use the keyboard to write the solution of the problem, which was one of the proper names that appeared in the question. The instructions stressed that the problems were to be answered as quickly and as accurately as possible, but emphasising accuracy over speed.

Before the presentation of the experimental materials there were three practice trials whose primary purpose was to familiarise participants with the procedure.

Results and Discussion

The mean correct response reading times for the first and the second displays, the mean correct response question answering times, and the percentage of correct responses for one- and two-model problems are shown in Table 2. The first display of one-model problems was read more quickly than the first display of two-model problems \(t(30)=4.86 \ P < .0001\]. Second display reading times did not show any reliable differences \(t(30)=1.19 \ P < 1\], neither did the question answering times \(t(30)=0.25 \ P < 1\]. However, one-model problems were responded to more accurately than two-model problems (Wilcoxon’s T; \(Z=1.98, \ N= 31, \ P < .05\).
TABLE 2
Experiments 2a and 2b

<table>
<thead>
<tr>
<th></th>
<th>One model (I)</th>
<th>Two models (II)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First display</td>
<td>17.6 (8.4)</td>
<td>24.6 (8.9)</td>
</tr>
<tr>
<td>Second display</td>
<td>12.8 (5.3)</td>
<td>14.1 (7.1)</td>
</tr>
<tr>
<td>Question times</td>
<td>5.8 (4.4)</td>
<td>6.0 (3.9)</td>
</tr>
<tr>
<td>Correct responses</td>
<td>87</td>
<td>75</td>
</tr>
<tr>
<td><strong>Nonspatial problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First display</td>
<td>12.7 (5.6)</td>
<td>17.7 (8.9)</td>
</tr>
<tr>
<td>Second display</td>
<td>7.9 (3.5)</td>
<td>8.0 (4.0)</td>
</tr>
<tr>
<td>Question times</td>
<td>3.9 (4.1)</td>
<td>4.3 (4.4)</td>
</tr>
<tr>
<td>Correct responses</td>
<td>92</td>
<td>83</td>
</tr>
</tbody>
</table>

Mean correct response reading times for the two displays and question answering times (in seconds), standard deviations (parentheses), as well as the percentage of correct responses to the problems of Experiments 2a and 2b.

These results replicate those of Experiments 1a and 1b by showing that two-model problems were harder than one-model problems. In addition, these results suggest that participants construct a mental model of the problem incrementally. When participants read the two first premises they seem to place the foundation for a representation that can contain either one model (problem I) or two models (problem II). It is harder to construct a two-model representation than a one-model representation. The processing costs that had occurred when participants read the two first premises still seem to be present when they read premises 3 and 4 (although the differences were not reliable), and when they develop the solution for the problem. In contrast, this pattern of data is difficult to explain by the inference-rule hypothesis, which predicted the opposite results.

**EXPERIMENT 2B**

In the previous experiment we have again shown that participants seem to create a mental model of the spatial layout that is described, and that they do so as soon as possible. In fact, we showed that they start to build a representation when they read the first two premises. In this experiment we investigate whether participants follow the same strategy with two-dimensional problems that are not spatial. In this experiment we used the same problems as those of Experiment 1b and the same presentation procedure as that of Experiment 2a.
Method

Participants. The participants were 27 undergraduate students from the University of La Laguna. They received course credit for participating. None of them had taken part in the previous experiments.

Design and Materials. The design and materials were the same as those of Experiment 1b.

Procedure. The procedure was the same as that of Experiment 2a.

Results and Discussion
The mean correct response reading times for the first and the second displays, the mean correct response question answering times, and the percentage of correct responses for one- and two-model problems are shown in Table 2. The first display of one-model problems was read more quickly than the first display of two-model problems \[ t(26)=3.86, \quad P < .001 \] No reliable differences were found between one-model problems and two-model problems either in second display reading times \[ t(26)=0.32, \quad P < 1 \], or in question answering times \[ t(26)=1.69, \quad P < 1 \]. Finally, responses to one-model problems were more accurate than responses to two-model problems (Wilcoxon’s T; \( Z=2.13 \), N=27, \( P < .05 \)).

Again, these results replicate those obtained in Experiments 1a and 2a with spatial descriptions and in Experiment 1b with nonspatial problems. Therefore, they suggest that participants reason by constructing mental models of the situations described in the texts, even when the situations are not spatial. The inference-rule hypothesis cannot account for these results.

GENERAL DISCUSSION
Taken together, the results of the two experiments suggest that participants solved the two-dimensional problems by building a mental model of the situations described in the texts. Moreover, they seem to build such representations incrementally by establishing a representational foundation as soon as possible which is updated with the incoming information. Another important finding of our experiments is that reasoning by building mental models of the situations is not necessarily restricted to problems that describe spatial relations. We have found similar results with problems that described spatial and nonspatial relations, showing in both cases that two-model problems were harder than one-model problems. However, as we pointed out, it could be still argued that mental models reasoning is circumscribed to the spatial domain, so nonspatial problems are translated or converted first into spatial problems in order to be solved. If this were the case, we should expect
that participants should be slower in solving nonspatial problems as compared to spatial problems. The data did not show this tendency. In contrast, participants were slower with spatial problems than with nonspatial problems. Schaecken, Johnson-Laird, and d’Ydewalle (in press, 1996; and Schaecken & Johnson-Laird, 1995) have recently reported similar results with temporal relations (e.g. a happens before b, b happens before c, etc.), finding that participants take longer to read the premises leading to multiple models.\footnote{This research is rather related to our own. We learned about it during the reviewing process of the present paper.}

In all of these experiments we failed to find evidence for the use of inference rules. Whether or not our results support the model theory of reasoning could be arguable. Rips (1994) has recently claimed that the experiments of Byrne and Johnson-Laird (1989) do not prove that the formal rules theories are wrong and the model theory is right. His argument relies on three objections (Rips, 1994). First, Rips considers that the instructions requesting the participants to form spatial arrays in solving the problems could bias the participants to use an “imaginal strategy” favouring the model theory predictions. We did not make use of such instructions in our experiments and we still found support for model-based strategies. Moreover, in our Experiments 1b and 2b, the problems are not represented in a two-dimensional Euclidean space because the two dimensions (study more or less, and copying from a given person) are nonspatial and essentially different from each other. Again, in these experiments we found support for model-based strategies. The second objection is that the presence of the irrelevant first premise (see Problem II) could sidetrack the search for a correct proof. In our opinion, the crucial difference between problems of type I and problems of type II is not the presence of the irrelevant premise but the need in any rule system for an additional inference in Problem I to determine the relation between the pair of items to which $D$ and $E$ are related. In our Experiment 2 (a and b) we found reliable differences in the first display between problems I and II when participants are still unaware of whether or not one of the premises would be irrelevant for the problem.

The third objection makes reference to the method of counting models used by Byrne and Johnson-Laird (1989). The claim is that in other tasks, such as syllogistic inference, models are considered genuinely different only if they support different conclusions (see, Johnson-Laird & Bara, 1984). Consequently, as they support the same conclusion, both problems I and II could be considered one-model problems. The vagueness in the method of counting models has also been pointed out by Bonatti (1994). As Johnson-Laird, Byrne, and Schaecken (1994) concede, the method of counting models in some domains, such as conditional reasoning, may be controversial because
participants may need to flesh out initial representations, and the extent of this fleshing-out process affects the number of models considered. However, in relational reasoning the number of models representing a given problem is easy to determine: it is the number of situations that could be consistent with the phrasing of the problem. In this sense, our Problem I is consistent with a single state of affairs while our Problem II could be consistent with two different states of affairs. The controversial point in our problems is not the number of models that a given problem yields, but whether or not participants represent one or two models, when confronted with problems of type II, to find the conclusion. It could be argued that some participants are not testing both models in Problem II when reaching the correct conclusion. Again, our second experiment solves the controversy. The reading times of the first display are greater in problems of type II, precisely when both models are needed for complete comprehension of the possible situations. It is possible that most participants do not scan both models in later stages because there is no need to do so to reach the conclusion, and that could be the reason why no differences in reading times were found in these later stages. However, participants seem to build both models while reading the first display.

Bonatti (1994) considers that the model theory is not an appropriate explanation for human reasoning (especially in the propositional domain). However, he admits that in some domains, such as the comprehension of texts that describe spatial relations, people could represent knowledge with nonpropositional structures. According to this argument, different mechanisms should govern the resolution of spatial and nonspatial problems; that is, formal rules for nonspatial problems, but a sort of fanatical mechanism for spatial problems. Our results suggest that participants did not follow different strategies to solve spatial and nonspatial problems. Evidence favouring the use of model representations in reasoning about relations has been obtained in the experiments described in this paper. Furthermore, Bauer and Johnson-Laird (1993) showed that the use of diagrams improves performance in propositional reasoning, by increasing the number of valid conclusions as well as the speed of finding a solution. If a specific mechanism governs the representation of spatial problems, no facilitation in propositional reasoning by means of diagrams should occur.

As well as providing theories of reasoning, the mental models theory also covers other aspects of cognition. The mental models theory has been proved successful as a framework for developing new hypotheses not only about human reasoning but also in the text comprehension field. There is a general agreement in the text comprehension literature that the end-product of the text comprehension process entails more than a propositional representation of a text (Garnham, 1981; Johnson-Laird, 1983; Sanford & Garrod, 1981; van Dijk & Kintsch, 1983). Participants seem to represent what the text is about, and not the structure of the text in which it is described, or the structure of the
sentences of that text. Moreover, in the comprehension field, it has been shown that participants construct representations (mental models) of the text that incorporate not only spatial but also nonspatial information (e.g., Carreiras, Garnham, Oakhill, & Cain, 1996; Gernsbacher, Goldsmith, & Robertson, 1992). Therefore, according to the mental models hypothesis it is not necessary to look for specific mechanisms to explain human reasoning and text comprehension, but the same general cognitive mechanisms underlie comprehension and human reasoning.

Manuscript received 19 December 1995
Revised manuscript received 24 October 1996

REFERENCES


